

## Red Clump Stars from the LAMOST data I: identification and distance

Junchen Wan<sup>1</sup>, Chao Liu<sup>1</sup>, Licai Deng<sup>1</sup>, Wenyuan Cui<sup>2</sup>, Yong Zhang<sup>3</sup>, Yonghui Hou<sup>3</sup>, Ming Yang<sup>1</sup>, Yue Wu<sup>1</sup>

<sup>1</sup>Key Laboratory of Optical Astronomy, National Astronomical Observatories, Chinese Academy of Sciences, 20A Datun Road, Beijing, 100012, China; junchenwan@bao.ac.cn

<sup>2</sup>Hebei Normal University, Shijiazhuang, China

<sup>3</sup>Nanjing Institute of Astronomical Optics & Technology, National Astronomical Observatories, Chinese Academy of Sciences, Nanjing 210042, China

**Abstract** We present a sample of about 120,000 red clump candidates selected from the LAMOST DR2 catalog based on the empirical distribution model in the effective temperature vs. surface gravity plane. Although, in general, red clump stars are considered as the standard candle, they do not exactly stay in a narrow range of absolute magnitude, but may extend to more than 1 magnitude depending on their initial mass. Consequently, conventional oversimplified distance estimations with assumption of fixed luminosity may lead to systematic bias related to the initial mass or the age, which may potentially affect the study of the evolution of the Galaxy with red clump stars. We therefore employ an isochrone-based method to estimate the absolute magnitude of red clump stars from their observed surface gravities, effective temperatures, and metallicities. We verify that the estimation well removes the systematics and provide an initial mass/age independent distance estimates with accuracy less than 10%.

**Key words:** stars: general—stars: horizontal-branch—stars: statistics—stars: distances—Galaxy: stellar content

### 1 INTRODUCTION

Red clump (RC) stars are metal-rich stars in the evolution phase of helium-core burning (Cassisi & Salaris 1997). They play important roles in the study of the Galactic disk because they are widespread in the thin disk, usually considered as the standard candle (Paczynski & Stanek 1998) and luminous (Girardi et al. 1998; Alves 2000; Groenewegen 2008) and prominent population in color magnitude diagram, which makes them be easily identified from multi-band photometry (López-Corredoira et al. 2002).

Identification of individual field RC stars, however, is not trivial because they do highly overlap with the RGB stars. Paczynski & Stanek (1998) used a Gaussian to model the distribution of magnitude for RC stars and a quadratic polynomial for RGB stars. Then the stars located in the Gaussian dominated region are very likely to be RC stars. This method was applied by Nataf et al. (2013) to select the red clump stars in the Galactic bulge. Recently, Bovy et al. (2014) employed a new method to identify the RC stars based on their distribution in color–effective temperature–surface gravity–metallicity and stellar evolution model. However, the authors only identify the primary RC stars and remove the secondary population to simplify the distance estimation.

Compared to the identification, the distance estimation of RC stars are relatively simple with the assumption that the absolute magnitude of RC stars is around a fixed value with a small dispersion (Paczynski & Stanek 1998; Girardi et al. 1998; Alves 2000; Groenewegen 2008; Liu et al. 2012).

However, the stellar evolution model demonstrates that the RC stars are not always stay in the same luminosity. They are separated into two subclasses: the primary RC stars, which have electronic-degenerate cores, and the second RC stars, which contain non-degenerate He-cores (Girardi 1999). In general, the primary RC stars are low-mass and hence older, while the secondary RC stars are massive and therefore younger than 1 Gyr. Most of the primary RC stars are generally brighter than the secondary RC stars, thus the former have smaller  $\log g$  than the later (see Stello et al. 2013). In the context of the evolution of the Galactic disk, we intend to obtain a sample of RC stars with a wide range of the age so that we can trace the evolution of the Galaxy from present day back to  $\sim 10$  Gyr. For instance, Salaris & Girardi (2002) fitted the distribution of the *Hipparcos* (Perryman et al. 1997) observed RC stars, including both the primary and secondary populations, in color-magnitude diagram with a stellar evolution model and derived the distribution of age in the solar neighborhood. Although keeping both the primary and secondary RC stars in the sample is important in the study of the disk evolution, the distance estimation may turn out to be more complicated since the RC stars would not be the standard candle any more in this case.

As the first paper of a series of works on the evolution of the Galactic disk with the RC stars from the LAMOST catalog, we firstly develop a new method of identification for both the primary and secondary RC stars and a normal approach of distance estimation to both populations. This paper is organized as below. In Section 2, we briefly introduce the LAMOST survey data and describe the empirical method of identification of RC stars. In Section 3, we develop the new method of the distance estimation. The performance of the estimates is then assessed in the same section. Finally, we further discuss the accuracy of our distance estimation in Section 4 and draw a brief conclusion in Section 5.

## 2 IDENTIFICATION OF RC STARS

In the section we identify the red clump stars from the LAMOST data using an empirical method.

### 2.1 The LAMOST data

The LAMOST, also called Guo Shou Jing telescope, is a 4-meter reflected Schmidt telescope with 4000 fibers on the 20-squared degree focal plane (Cui et al. 2012; Zhao et al. 2012). The LAMOST survey will finally observe more than 5 million low resolution stellar spectra after its 5-year survey (Deng et al. 2014; Liu et al. 2014). According to Yao et al. (2012), the site of LAMOST prefers to observe the winter sky, which covers the Galactic anti-center region. Therefore, there will be lots of disk populations located in the region to be observed by the LAMOST. In this work, we adopt the derived  $T_{\text{eff}}$  directly from the LAMOST pipeline (Wu et al. 2011a; Wu et al. 2011b; Wu et al. 2014; Luo et al. in prep) and the estimated  $\log g$  from Liu et al. (2015), who estimate  $\log g$  using support vector regression with the training dataset from the Kepler asteroseismic  $\log g$ . The uncertainty of the  $\log g$  estimates is only about 0.1 dex, which is a factor of 2 better than that from the LAMOST pipeline, for the metal-rich ( $[\text{Fe}/\text{H}] > -0.6$ ) giant stars, including the RC stars.

We select the stars with  $\log g$  between 0.9 and 3.5,  $[\text{Fe}/\text{H}]$  between -0.6 and 0.4, and  $T_{\text{eff}}$  between 3600 and 6000 K from the LAMOST DR2 catalog, which widely cover the RC region. We also remove all spectra with signal-to-noise ratio lower than 10 and finally obtain 279,423 stars.

### 2.2 Identification of RC stars

It is expected that lots of red clump stars in the disk population will be sampled by LAMOST survey. Indeed, Fig 1 shows that the RC stars from LAMOST DR2 catalog are prominent in  $T_{\text{eff}}$  vs.  $\log g$  plane. In this section, we establish an approach of identification of RC stars from the LAMOST data.

Paczyński & Stanek (1998) used a Gaussian and quadratic polynomial to model the distribution of the magnitude for the RC and RGB stars, respectively. We expand this method into two dimensions in  $T_{\text{eff}}$  vs.  $\log g$  plane.

First, we empirically build a 2-D distribution model for the RGB stars in  $T_{\text{eff}}$  vs.  $\log g$  plane with various metallicity bins ( $[\text{Fe}/\text{H}] = (-0.6, -0.3), (-0.3, 0.0), \text{ and } (0.0, 0.4)$ ). We mask the region between  $\log g = 2.0$  and  $3.0$  to avoid the RC stars and fit the distribution of the rest RGB stars with the following empirical model,

$$N_{\text{RGB}}(T_{\text{eff}}, \log g) = N(\log g) * \exp\left(-\frac{(T_{\text{eff}} - T(\log g))^2}{\sigma^2(\log g)}\right) \quad (1)$$

where  $N(\log g)$  and  $\sigma^2(\log g)$  are three smoothing cubic spline functions with different metallicity bins ( $[\text{Fe}/\text{H}] = (-0.6, -0.3), (-0.3, 0.0), \text{ and } (0.0, 0.4)$ ), respectively; and

$$T(\log g) = \begin{cases} 150.2\log g^3 - 946.3\log g^2 + 2417\log g + 2188 & [\text{Fe}/\text{H}] \in (-0.6, -0.3) \\ 49.8\log g^3 - 374.1\log g^2 + 1435\log g + 2632 & [\text{Fe}/\text{H}] \in (-0.3, 0.0) \\ 81.65\log g^3 - 590.7\log g^2 + 1921\log g + 2107 & [\text{Fe}/\text{H}] \in (0.0, 0.4) \end{cases} \quad (2)$$

Fig 2 shows the best fit curves for the terms in Eq 1- 2. From left to right, the range of metallicity is  $(-0.6, -0.3), (-0.3, 0.0), \text{ and } (0.0, 0.4)$ , respectively.  $N(\log g)$  and  $\sigma^2(\log g)$  are the best fit smoothing spline functions shown in the top panels and bottom panels of Fig 2, respectively. The rest panels of Fig 2 show the best fit polynomials for the terms in Eq 2.

Providing that the RGB stars are smoothly distributed along  $\log g$ , we can *interpolate* the stellar distribution of the RGB stars between  $\log g = 2.0$  and  $3.0$  with the best fit model shown in Eqs 1- 2. The middle panels of Fig 3 shows the distribution of RGB stars according to our model in  $T_{\text{eff}}$  vs.  $\log g$  plane for various  $[\text{Fe}/\text{H}]$  bins in different rows.

Second, we subtract the smoothing distribution model of RGB stars between  $\log g = 2$  and  $3$  and the residuals are mostly contributed by the RC stars, as shown in the right panels of Fig 3, in which the contours of 68% (red) and 95% (yellow) completeness of the residual distribution are overlapped. Compromise has to be made between the completeness for both primary and secondary RC stars and the fraction of the contamination from the RGB stars. We find that the 68% completeness contour cannot cover the most region of the secondary RC stars, therefore, we select the 95% completeness, as the recommended selection criterion of the RC stars. With this empirical distribution of RC stars, other users can freely adjust the criterion of selection to identify different sample of the RC stars to meet their specific requirements. It is noted that there are a few fractional areas far way from the empirical RC star region are also within the 95% completeness level. They may be artificial features because the data is too sparse in these regions. We then manually exclude three artificial areas with  $T_{\text{eff}} > 5200 \text{ K}$ ,  $\log g < 2.1 \text{ dex}$  and  $\log g > 2.9 \text{ dex}$ . Moreover, the bottom-right corner of the 95% completeness level in the bottom-right panel is apparently contributed by the red bump stars rather than the clump stars. Therefore, a fourth cut at  $\log g < 0.0016T_{\text{eff}} - 4.7170$  is added to remove the contamination from the bump stars. These additional data cuts are shown as white lines in the right panels of Fig 3. Finally, we find 118,711 RC candidates with the refined 95% completeness level.

Although most of the stars located in the refined 95%-level region are RC stars, some contaminations may also included. Assuming that the residual distribution in  $T_{\text{eff}}$  vs.  $\log g$  planes of Fig 3 are the distributions of the true RC stars, we can give the percentage of the true RC stars by dividing the residual distributions with the full distributions, which contain both RC and RGB stars, although we cannot identify the individual RC stars. Fig 4 shows the percentage in  $T_{\text{eff}}$  vs.  $\log g$  planes for metallicity bin  $(-0.6, -0.3), (-0.3, 0.0), \text{ and } (0.0, 0.4)$  from left to right, respectively. It shows that the fraction of the true RC stars are mostly larger than 75%, even for some regions of the secondary RC stars, in the candidate catalog. The contour level of the 68% and 95% completeness, the numbers of the RC candidates under such completeness levels, and the total fraction of the true RC stars are listed in Table 1.

Compared with the method provided by Bovy et al. (2014), our method does not depend on the stellar model but only on the specific observations. Moreover, although the number of the secondary RC are less than that of the primary RC, we can still discriminate them from the background RGB stars with a fraction of 75%~85%, as shown in Fig 4. Therefore, the identification of RC stars in this work is

suitable for both primary and secondary RC stars, ensuring that the study of the evolution of the Milky Way can be extended from  $<1$  Gyr to around 10 Gyr based on this sample.

### 3 DISTANCE ESTIMATIONS

Most of the RC stars are located in the Galactic disk and their apparent magnitudes and color indices are significantly affected by the interstellar extinction. Therefore, the distance and the reddening have to be determined simultaneously. In the next section, we introduce a likelihood method to determine these quantities, and then we apply this method with fixed and various absolute magnitude, respectively, in the following two sections.

#### 3.1 A likelihood method to estimate distance and reddening

The first step to estimate the distance of RC stars is to estimate the reddening from the observed color index. The LAMOST spectra do not have a perfect optical multi-band photometry by now. Its input catalog is a combination of UCAC4 (Zacharias et al. 2013), PanSTARRS1 (Tonry et al. 2012), SDSS (Ahn et al. 2014), and Xuyi Galactic anti-center survey (Yuan et al. 2015). Although all of these source catalogs contains  $g$ ,  $r$ , and  $i$  bands, they are still not well calibrated with each other, yet. Therefore, in this stage, we use the 2MASS catalog (Skrutskie et al. 2006) as the input catalog to derive the reddening and distance for RC stars.

The likelihood of  $E(J - K)$  for a RC star given the observed  $J - K$  and the intrinsic color index of RC stars can be written as:

$$Pr(E(J-K)|J-K, \sigma_{J-K}, (J-K)_{RC}, \sigma_{RC, J-K}) \sim \exp\left(-\frac{(E(J-K) - ((J-K) - (J-K)_{RC}))^2}{2(\sigma_{J-K}^2 + \sigma_{RC, J-K}^2)}\right), \quad (3)$$

where  $(J-K)_{RC}$  is the intrinsic color index of RC stars,  $\sigma_{RC, J-K}$  the intrinsic color index dispersion of RC stars, and  $\sigma_{J-K}$  the measurement error of the observed  $J-K$ . To convert the reddening in  $J-K$  to the extinction in  $K$  band, we adopt Indebetouw et al. (2005) that

$$A_K = 0.67E(J-K). \quad (4)$$

Then, the likelihood of the distance module,  $DM$ , for a RC star given the observed  $K$  magnitude and the fixed absolute magnitude  $M_K$  can be written as:

$$Pr(DM|K, \sigma_K, A_K, M_K, \sigma_{M_K}) \sim \exp\left(-\frac{(DM - (K - A_K - M_K + 5))^2}{2(\sigma_K^2 + \sigma_{M_K}^2)}\right), \quad (5)$$

where  $M_K$  is the fixed absolute magnitude in  $K$  band for RC stars,  $\sigma_{M_K}$  the intrinsic dispersion of absolute magnitude in  $K$  band for RC stars, and  $\sigma_K$  is the measurement error of the apparent  $K$  magnitude.

#### 3.2 The fixed absolute magnitude and the intrinsic color index of the RC stars

Combined with Eqs. 3- 5, the likelihood of both  $E(J - K)$  and  $DM$  for RC stars can be derived. The last ingredient needed to put in is the absolute magnitude, intrinsic color index, and their intrinsic dispersions for RC stars. Although some literatures provided the absolute magnitude in  $K$  band and the intrinsic absolute magnitude in  $J - K$  for RC stars (Alves 2000; Groenewegen 2008; Zasowski et al. 2013 etc.), the intrinsic dispersions for both quantities, which are necessary in our likelihood method, are not self-consistently provided. Therefore, we estimate the absolute magnitude in  $K$  band, the intrinsic color index in  $J - K$ , and their intrinsic dispersions with the *Hipparcos* data.

*Hipparcos* catalog provides parallaxes for more than 100,000 bright stars, thousands of which located in the region of red clump in HR diagrams. In order to estimate the absolute magnitude of RC stars, we need to correct the reddening as the first step. Bailer-Jones (2011) estimated the extinction parameters for about 47,000 *Hipparcos* stars, sparsely distributed in the sky. We extend the extinction

to all *Hipparcos* stars using spatial interpolation. For a star of interest, we select all stars with reddening parameters derived by Bailer-Jones (2011) within a 10-degree-radius circle and 20 pc in distance around it. Then we assign the median reddening value for all selected stars as the reddening value for the star of interest. In order to ensure the accuracy of the absolute magnitude, we select the stars with error of parallax smaller than 20% and errors of 2MASS photometry smaller than 0.5 mag. Fig 5 shows the  $J - K$  vs.  $M_K$  diagrams without (top-left panel) and with dereddening (middle-left panel) for about 900 giant stars with  $M_K < 0$  mag.

We adopt the empirical model of the distribution in  $M_K$  from Paczyński & Stanek (1998), which has the following form:

$$F = a + b(M_K - c)^2 + d \exp\left(-\frac{(M_K - e)^2}{2f^2}\right), \quad (6)$$

where  $a, b, c, d, e$ , and  $f$  are the free parameters. The quadratic polynomial in Eq (6) models the stellar distribution of the RGB stars and the Gaussian term models the RC stars.

Similarly, we use this model for the marginalized distribution of color index:

$$F' = a' + b'(J - K - c')^2 + d' \exp\left(-\frac{(J - K - e')^2}{2f'^2}\right). \quad (7)$$

We fit the models for the absolute magnitude and color index without and with the dereddening, respectively. The top-right panel of Fig 5 shows the best fit for the marginalized reddened absolute magnitude with Eq (6). The middle-right panel shows the best fit model for the marginalized dereddened absolute magnitude. And the bottom panel shows the best fit for the reddened (dashed line) and dereddened (solid line) color index  $J - K$ , respectively. Table 2 lists the best fit absolute magnitude and intrinsic color index and their dispersions. It shows that the dereddend absolute magnitude and intrinsic color index are brighter and bluer than the reddened values by about 2%, respectively. In this work we adopt the dereddened  $M_K$  and  $J - K$  as the standard value in the estimation of the distances for RC stars.

### 3.3 The $M_K$ based on the synthetic isochrones

The assumption that the RC stars have a fixed magnitude can only work for the primary RC stars, which are mostly composed of the relatively older RC stars compared with the secondary ones. When we want to trace the evolution of the Galactic disk with RC stars, we cannot only use the primary RC stars and ignore the secondary RC stars. Therefore, we need to improve the approach to estimate the distance of RC stars so that the secondary RC stars are also taken into account.

We turn to use a special isochrone fitting process to estimate the absolute magnitude for all kinds of RC stars. After quickly reviewing the isochrones, we realize that the absolute magnitude of RC stars is a function of  $\log g$ ,  $[\text{Fe}/\text{H}]$ , and  $T_{\text{eff}}$ . Panel (a) of Fig 6 shows the synthetic RC stars from PARSEC library (Bressan et al. 2012) in  $\log g$  vs.  $M_K$  plane. It shows that  $M_K$  is tightly dependent of  $\log g$ . Further separations of the data into different initial stellar masses are shown in the right panels. We find that the RC stars with stellar masses of  $0.8 \sim 1.1 M_{\odot}$  (panel (b) in Fig 6) mostly concentrate within  $M_K = -1.4 \sim -1.6$  mag, while the  $M_K$  for the RC stars with initial mass of  $1.1 \sim 1.4 M_{\odot}$  moves up to  $-1.6 \sim -1.8$  mag (panel (c)). Then the  $M_K$  for the RC stars with initial mass at  $1.4 \sim 1.7 M_{\odot}$  move back to the range of  $-1.5 \sim -1.7$  mag (panel (d)). For the RC stars with  $1.7 \sim 2.0 M_{\odot}$ ,  $M_K$  intensively extends from  $\sim -1$  to more than  $-2$  mag (panel (e)). To further investigate how the  $M_K$  varies, we separate the synthetic stars into two groups at  $[\text{Fe}/\text{H}] = -0.3$ . For the RC stars with  $[\text{Fe}/\text{H}] > -0.3$  dex, the relation between  $M_K$  and  $\log g$  can be empirically modeled with a quadratic polynomial (see the related panels in Fig 7):

$$M_K = P_1 \log g^2 + P_2 \log g + P_3 \quad (8)$$

The best fit coefficients  $P_i$  ( $i = 1, 2, 3$ ) are listed in Table 3. For the stars with  $[\text{Fe}/\text{H}] < -0.3$  dex, when  $\log g < 2.45$ , the  $M_K$  is no longer a function of  $\log g$  (see the related panels in Fig 7). Fig 8 shows that



for these stars, the absolute magnitude is roughly a linear function of the effective temperature. Then we have the following more complicated relation:

$$M_K = \begin{cases} P_1 T_{\text{eff}} + P_2 & \log g < 2.45 \\ P_1 \log g^2 + P_2 \log g + P_3 & \log g > 2.45 \end{cases} \quad (9)$$

Table 4 shows the best fit coefficients of  $P_i$  ( $i = 1, 2, 3$ ). It is noted that for both groups of metallicity, the synthetic data (red dots) showing in Figs 7 and 8 are not exactly located on a narrow line, but spread out with various dispersions. We then measure the dispersions of the residuals of the  $M_K$  for the synthetic RC stars with respect to the best fit models at different bins of metallicity and show them in the column of  $\sigma_{M_K}$  in Tables 3 and 4.

Then, for each observed RC star, we firstly derive the  $M_K$  and  $\sigma_{M_K}$  from its  $T_{\text{eff}}$ ,  $\log g$ , and  $[\text{Fe}/\text{H}]$ , and bring them back to Eq 5 to derive the likelihood of the distance module. Because the intrinsic color index of the primary and secondary RC stars are quite similar, we adopt this value derived from section 3.2.

### 3.4 Performance assessment

Before applying the improved distance estimation to the LAMOST data, we assess the performance of the improved  $M_K$  model demonstrated in section 3.3.

We arbitrarily select 1000 points from the synthetic dataset and add random Gaussian errors to the true values of  $T_{\text{eff}}$ ,  $\log g$ , and  $[\text{Fe}/\text{H}]$ . For each synthetic RC star, we create 20 mock stars with different random errors. In total, we create 20,000 mock stars with errors. Then, we derive the absolute magnitude for the mock data based on the method mentioned in Section 3.3. We totally create 9 sets of mock dataset with various measurement errors of  $\log g$ ,  $T_{\text{eff}}$ , and  $[\text{Fe}/\text{H}]$ . In the first three mock datasets, we create the errors with  $\sigma_{\log g} = 0.1, 0.2$ , and  $0.3$  dex, respectively,  $\sigma_{T_{\text{eff}}} = 120$  K, and  $\sigma_{[\text{Fe}/\text{H}]} = 0.1$  dex. The second three mock datasets are created with random errors of  $\sigma_{T_{\text{eff}}} = 120, 150$ , and  $200$  K, respectively,  $\sigma_{\log g} = 0.1$  dex, and  $\sigma_{[\text{Fe}/\text{H}]} = 0.1$  dex. The last three mock datasets have  $\sigma_{[\text{Fe}/\text{H}]} = 0.1, 0.2$ , and  $0.3$  dex, respectively, with  $\sigma_{\log g} = 0.1$  dex, and  $\sigma_{T_{\text{eff}}} = 120$  K. We compare the derived absolute magnitudes with the true values in the 9 mock samples. The residuals of the derived absolute magnitudes as functions of the errors of the stellar parameters are shown in Fig 9.

The left panels of Fig 9 show the residuals of  $M_K$ , denoted as  $\delta M_K$ , as a function of the errors of  $\log g$  at  $\sigma_{\log g} = 0.1, 0.2, 0.3$  dex from top to bottom, respectively. It shows that the larger the uncertainty of  $\log g$ , the larger the errors in the absolute magnitude. When the random errors of  $\log g$  are larger, the derived  $M_K$  seems more overestimated. However, slightly increasing the uncertainties of  $[\text{Fe}/\text{H}]$  (the middle panels) and  $T_{\text{eff}}$  (the right panels) would not significantly add the uncertainties of the  $M_K$  estimates. Fig 10 shows the standard deviation of  $\delta M_K$ , in terms of the  $\sigma$  of the best fit Gaussian with the histogram of  $\delta M_K$ , as functions of the uncertainties of  $\log g$  (the left panel),  $[\text{Fe}/\text{H}]$  (the middle panel), and  $T_{\text{eff}}$  (the right panel), respectively. Again, this figure shows that the accuracy of the  $M_K$  estimates relies mostly on the accuracy of  $\log g$ , rather than that of  $[\text{Fe}/\text{H}]$  and  $T_{\text{eff}}$ . Therefore, the accurate  $\log g$  calibrated with the asteroseismic  $\log g$  from Liu et al. (2015) is very important in the high accuracy distance estimation.

Figure 10 also shows that with the typical uncertainty of 0.1 dex in  $\log g$ , the uncertainty of the derived absolute magnitude is better than 0.2 mag, corresponding to  $\sim 10\%$  in distance.

## 4 DISCUSSIONS

### 4.1 Comparison between two absolute magnitude models

In section 3, we discuss two approaches to estimate the absolute magnitude for RC stars. It is worthy to directly compare the distance estimates based on the two different methods. Fig 11 shows the difference of the distance estimates for the LAMOST RC stars between the fixed absolute magnitude-based and the isochrone-based method as a function of  $\log g$ . It is seen that the fixed magnitude-based method tends

to underestimate the RC stars with smaller  $\log g$  and significantly overestimate those with larger  $\log g$ . The overestimation in large  $\log g$  data is because that the fixed absolute magnitude are dominated by the primary RC stars and hence may not be suitable for the secondary RC stars. The underestimation when  $\log g < 2.7$  dex is likely because the slightly inconsistency between the  $M_K$  used in the synthetic isochrones and the one derived in section 3.2. In Fig 11, we find that the systematic bias can shift by more than 20% in large  $\log g$  given a fixed absolute magnitude. When  $\log g < 2.3$  dex, the isochrone-based method may not give reliable  $M_K$  estimates for RC stars since this is very close to the boundary of the isochrone data (see Fig 8). Therefore the errors increase in this region, as shown in Fig 11.

#### 4.2 External uncertainty of the distance estimation

In Section 3.4, we use the mock data created from the synthetic data to test the performance of the distance estimation based on the isochrones. This, however, can only give the internal error but not the external one. We then cross-identify the RC stars from the LAMOST data with *Hipparcos* catalog. Unfortunately, we only obtain less than 10 common RC stars with parallax error less than 50%. Because most of these RC stars suffers larger uncertainty in parallax, they cannot be treated as the standard stars to investigate the external error of the isochrone-based distance estimates. To solve this issue, we need to wait the coming *Gaia* data (Bailer-Jones 2009), which will release the first catalog in 2016.

### 5 CONCLUSIONS

In this work, we set up an empirical model in  $T_{\text{eff}}$  vs.  $\log g$  plane to identify the RC stars in the LAMOST DR2 data. The employed approach identifies not only the primary, but also the secondary RC stars. This will be very helpful for the study of the evolution of the Milky Way, because the range of the RC stars can be extended from  $<1$  Gyr (contributed by the secondary RC stars) to 10 Gyr (contributed by the primary RC stars). Finally, we identify 118,711 RC stars from the LAMOST DR2 data with the 95% completeness level. The sample of the selected RC stars may be contaminated by the RGB stars with the fraction of about 20%.

After the identification of the RC stars, we develop two different approaches to estimate their distances as well as the interstellar extinction. The first one is based on the fixed absolute magnitude and intrinsic color index. The accuracy of the fixed absolute magnitude-based approach relies on the accuracy of the determination of the absolute magnitude and the dispersion of the absolute magnitude. Consequently, we revisit the absolute magnitude and its dispersion for RC stars in  $K$  band. We adopt the similar empirical model introduced by Paczyński & Stanek (1998), but take into account the interstellar extinction for the *Hipparcos* RC stars. Although the extinction is small, it leads to the underestimation of the absolute magnitude by about 2%.

Although this method is sufficiently accurate for most of the primary RC stars, it is not suitable for the secondary RC stars, which significantly vary in  $\log g$  and hence also in the absolute magnitude. We then develop the second approach considering both types of RC stars based on the isochrones. With the empirical model, we associate the absolute magnitude of a RC star with its  $[\text{Fe}/\text{H}]$ ,  $\log g$ , and  $T_{\text{eff}}$ . The more delicate model can reduce the uncertainty of the distance estimates to 10% for almost all types of RC stars given the error of  $\log g$  at around 0.1 dex.

This RC sample well samples the Galactic disk, particularly in the outer disk, allowing us to map the structure, kinematics, and evolution of the Galactic disk in future works.

**Acknowledgements** This work is supported by the Strategic Priority Research Program “The Emergence of Cosmological Structures” of the Chinese Academy of Sciences, Grant No. XDB09000000 and the National Key Basic Research Program of China 2014CB845700. CL acknowledges the National Science Foundation of China (NSFC) under grants 11373032, 11333003 and U1231119. Guoshoujing Telescope (the Large Sky Area Multi-Object Fiber Spectroscopic Telescope LAMOST) is a National Major Scientific Project built by the Chinese Academy of Sciences. Funding for the project has been provided by the National Development and Reform Commission. LAMOST is operated and managed by the National Astronomical Observatories, Chinese Academy of Sciences.

## References

- Ahn, C. P., Alexandroff, R., Allende P., et al. 2014, *ApJS*, 211, 16
- Alves, D. R., 2000, *ApJ*, 539, 732
- Bailer-Jones, Coryn A. L., 2009, *IAU Symposium*, 254, 475
- Bressan, A., Marigo, P., Girardi, L., et al. 2012, *MNRAS*, 427, 127
- Bailer-Jones, C. A. L., 2011, *MNRAS*, 411, 435
- Bovy, J., Nidever, D. L., Rix, H.-W., et al. 2014, *ApJ*, 790, 21
- Cabrera-Lavers, A., Hammersley, P. L., González-Fernández, C., et al. 2007, *A&A*, 465, 825
- Cassisi, S., & Salaris, M. 1997, *MNRAS*, 285, 593
- Cui, X.-Q., Zhao, Y.-H., Chu, Y.-Q., et al., 2012, *RAA*, 12, 1197
- Deng, L.-C., Newberg, H. J., Liu, C., et al. 2012, *RAA*, 12, 735
- Girardi, L., Groenewegen, M. A. T., Weiss, A., et al. 1998, *MNRAS*, 301, 149
- Girardi, L. 1999, *MNRAS*, 308, 818
- Girardi, L., 2000, *IAU Joint Discussion*, 13, 10
- Groenewegen, M., A.T. 2008, *A&A*, 488, 935
- Indebetouw, R., Mathis, J. S., Babler, B. L., et al. 2005, *ApJ*, 619, 931
- Liu, C., Xue, X.-X., Fang, M. et al. 2012, *ApJ*, 753, 24
- Liu, X.-W., Yuan, H.-B., Huo, Z.-Y., et al., 2014, *Proceeding of IAUS 298*, Felzing, S., Zhao, G., & Walton, N. A. Eds., Cambridge University Press, 310 (arXiv: 1306.5376)
- Liu, C., Fang, M., Wu, Y., et al. 2015, *ApJ* in press, arXiv:1411.0235
- López-Corredoira, M., Cabrera-Lavers, A., Garzón, F., et al. 2002, *A&A*, 394, 883
- Nataf, D. M., Gould, A., Fouqué, P., et al. 2013, *ApJ*, 769, 88
- Paczynski, B. & Stanek, K., 1998, *ApJ*, 494, 219
- Perryman, M. A. C., Lindegren, L., Kovalevsky, J., et al. 1997, *A&A*, 323, L49
- Salaris, M. & Girardi, L., 2002, *MNRAS*, 337, 332
- Skrutskie, M. F., Cutri, R. M., Stiening, R., et al. 2006, *ApJ*, 131, 1163
- Stanek, K., 1996, *ApJ*, 460, 37
- Stello, D., Huber, D., Bedding, T. R., et al. 2013, *ApJ*, 765, L41
- Tonry, J. L., Stubbs, C. W., Lykke, K. R., et al. 2012, *ApJ*, 750, 14
- Wu, Y., Singh, H. P., Prugniel, P., 2011a, *A&A*, 525, 71
- Wu, Y., Luo, A., Li, H., et al. 2011b, *RAA*, 11, 924
- Wu, Y., Luo, A., Du, B., et al. 2014, arXiv:1407.1980
- Yao, S., Liu, C., Zhang H.-T., et al. 2012, *RAA*, 12, 772
- Yuan H.-B., Liu X.-W., Huo, Z.-Y., et al. 2015, *MNRAS*, 448, 855
- Zacharias, N., Finch, C. T., Girard, T. M., et al. 2013, *ApJ*, 145, 14
- Zasowski, G., Johnson, J. A., Frinchaboy, P. M., et al. 2013, *AJ*, 146, 81
- Zhao, G., Zhao, Y.-H., Chu Y.-H., et al. 2012, *RAA*, 12, 723



**Table 1** The location of the RC candidates in the right panel in Fig 3

[Fe/H]	Completeness level	Contour level	$N_{RC}$	Total ratio of RC stars
(-0.6,-0.3)	68%	51.39	36099	89.68%
	95%	15.46	54890	80.47%
(-0.3,0.0)	68%	42.26	31553	83.32%
	95%	10.46	48064	76.29%
(0.0,0.4)	68%	12.12	11291	84.87%
	95%	3.01	15757	81.13%

Notes:

Contour level: the contour level in Fig 3 corresponding to 68% or 95% completeness.

 $N_{RC}$ : number of RC candidates enclosed in the contour. (with the addition cut around the edges, see white lines in the right panels of Fig 3)

Ratio of RC stars: the total fraction of RC stars to the all metal-rich giant stars within the contour level.

**Table 2** The derived absolute magnitude and intrinsic color index of RC stars.

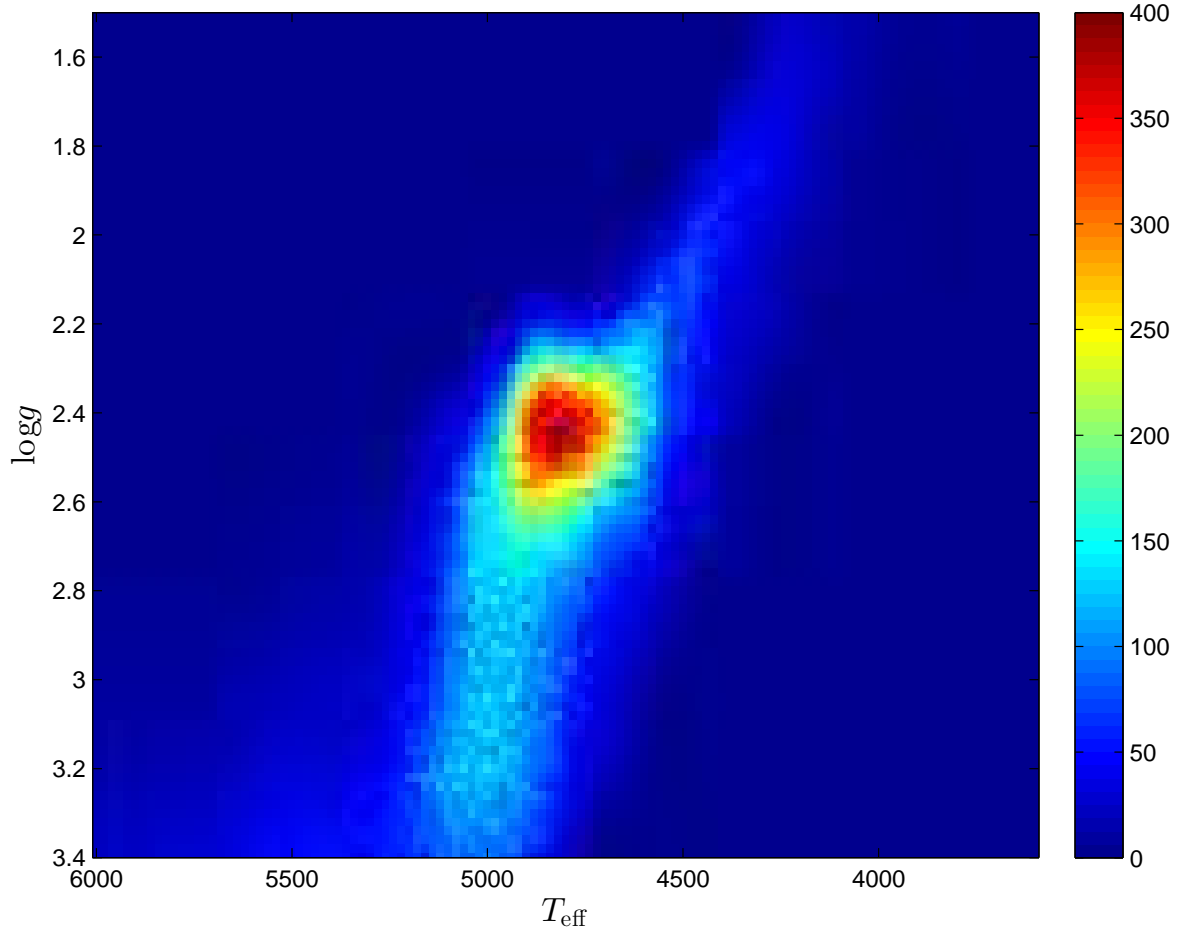
	$M_K$ mag	$\sigma_{M_K}$ mag	$J - K$ mag	$\sigma_{J-K}$ mag
reddened	$-1.529 \pm 0.003$	$0.075 \pm 0.003$	$0.681 \pm 0.003$	$0.104 \pm 0.005$
dereddened	$-1.549 \pm 0.003$	$0.076 \pm 0.003$	$0.658 \pm 0.001$	$0.072 \pm 0.001$

**Table 3** The coefficients at different [Fe/H] bins for the  $M_K$  model of metal-rich RC stars

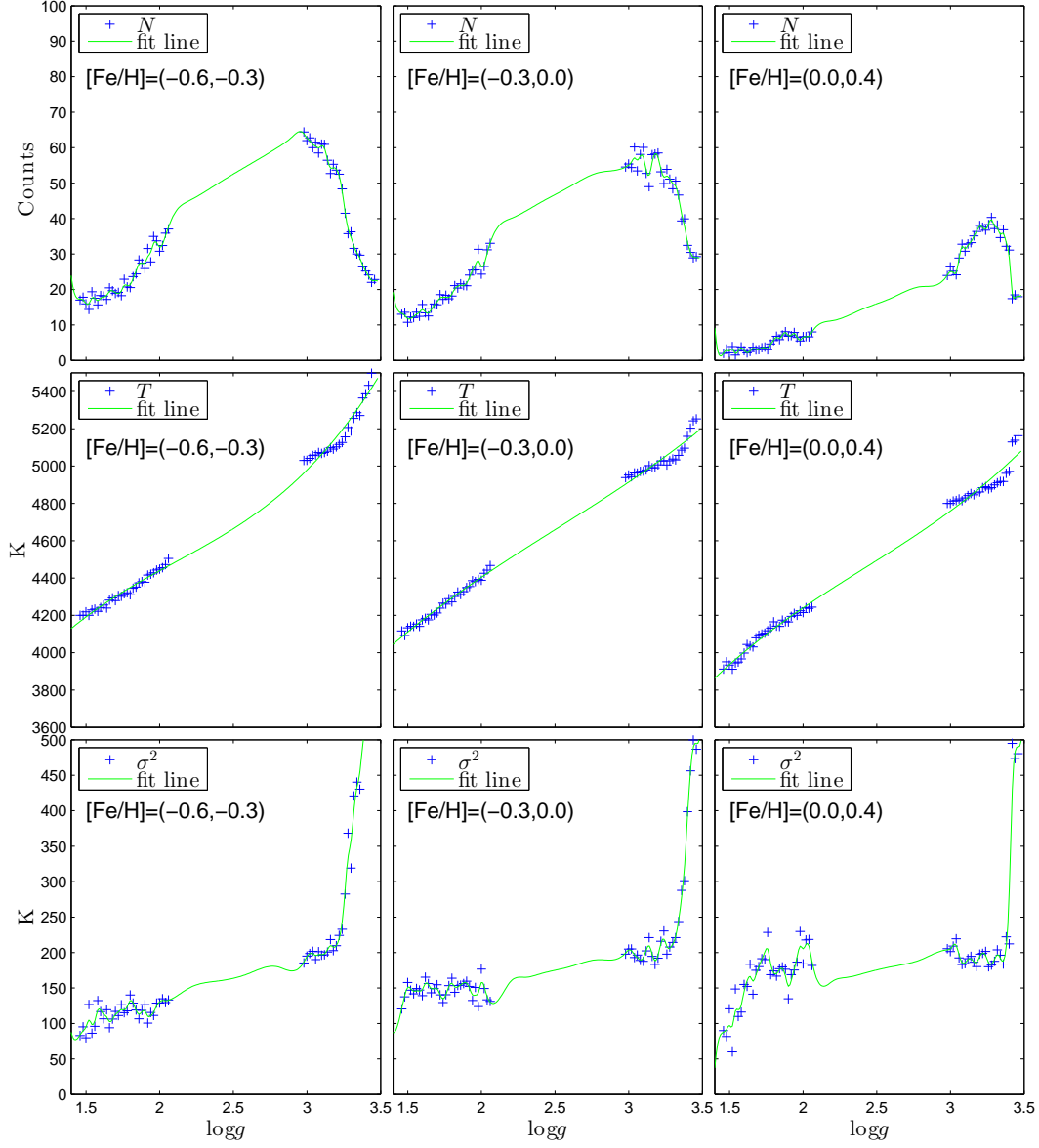
[Fe/H]	$M_K = P_1 \log^2 g + P_2 \log g + P_3$			
	$P_1$	$P_2$	$P_3$	$\sigma_{M_K}$
(-0.3,-0.2)	6.72	-33.94	41.15	0.12
(-0.2,-0.1)	6.67	-33.55	40.49	0.10
(-0.1,0.0)	6.21	-31.25	37.64	0.11
(0.0,0.1)	6.17	-31.00	37.26	0.10
(0.1,0.2)	6.19	-31.05	37.26	0.09

**Table 4** The coefficients at different [Fe/H] bins for the  $M_K$  model of metal-poor RC stars

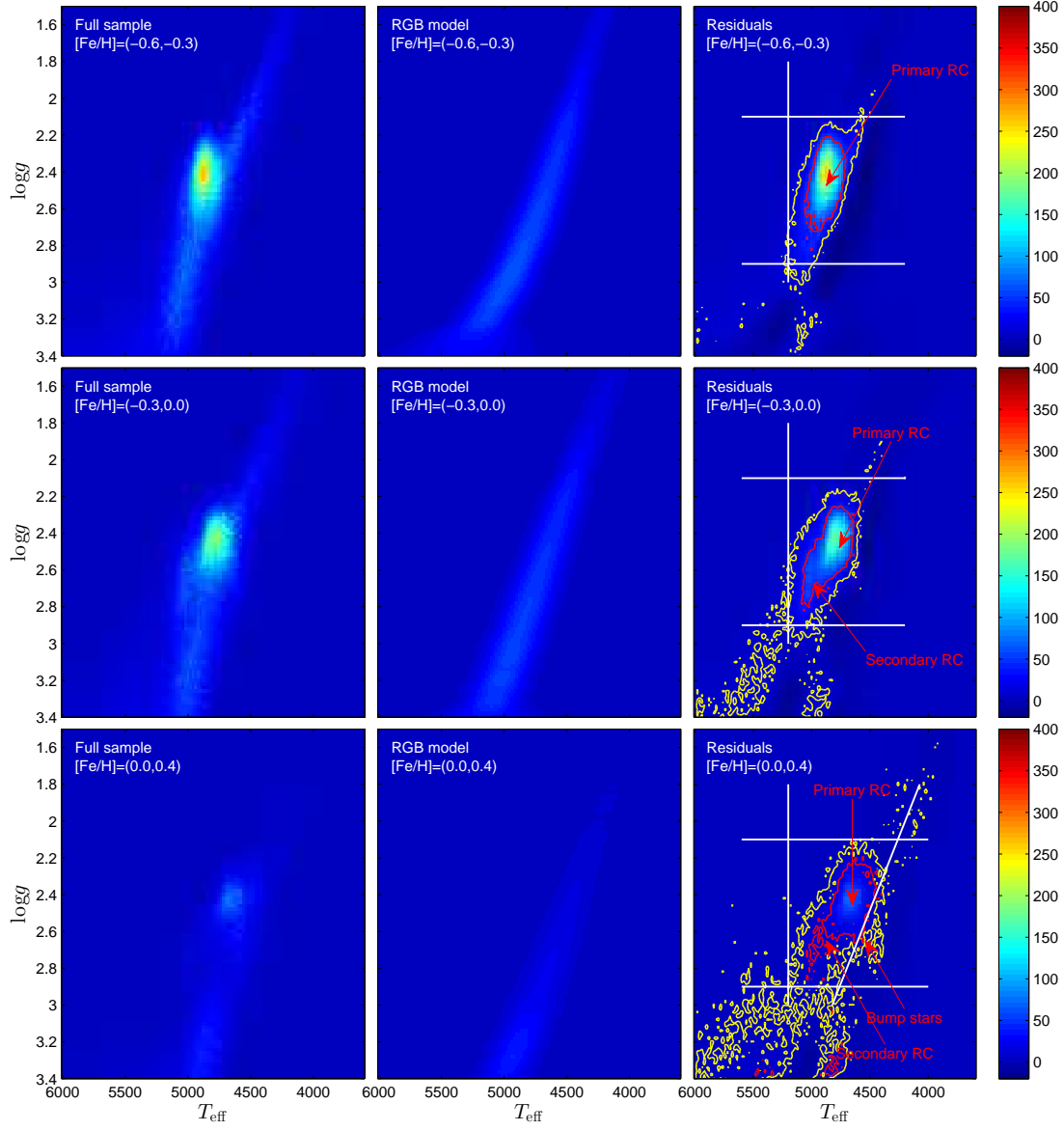
[Fe/H]	$\log g < 2.45$			$\log g > 2.45$			
	$M_K = P_1 T_{\text{eff}} + P_2$			$M_K = P_1 \log^2 g + P_2 \log g + P_3$			
	$P_1$	$P_2$	$\sigma_{M_K}$	$P_1$	$P_2$	$P_3$	$\sigma_{M_K}$
(-0.6,-0.5)	0.0024	-13.65	0.15	7.87	-39.98	49.06	0.09
(-0.5,-0.4)	0.0026	-14.50	0.14	6.28	-31.74	38.41	0.08
(-0.4,-0.3)	0.0027	-14.73	0.14	6.54	-33.09	40.16	0.08



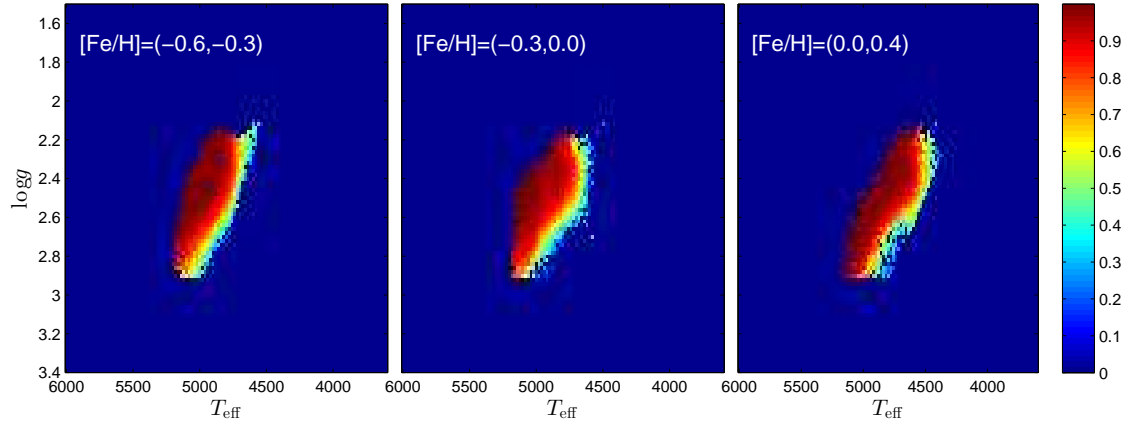
**Fig. 1** The observed distribution of the metal-rich ( $[\text{Fe}/\text{H}] > -0.6$ ) giant branch stars from LAMOST DR2 in  $\log g$  vs.  $T_{\text{eff}}$  plane. The bin size is  $\Delta \log g = 0.02$  and  $\Delta T_{\text{eff}} = 20\text{K}$ .



**Fig. 2** Fitting the coefficients in Eq 1. The solid lines are the best fit curve, while the cross symbols are measured from RGB stars with  $\log g < 2$  and  $> 3$ . From left to right, the metallicity bins are  $(-0.6, -0.3)$ ,  $(-0.3, 0.0)$ , and  $(0.0, 0.4)$ , respectively. The top panels show the best fit smoothing splines for  $N$ , the bottom panels show the best fit splines for  $\sigma^2$ . The rest panels show the best fit polynomials, which coefficients are appeared in Eq 2.

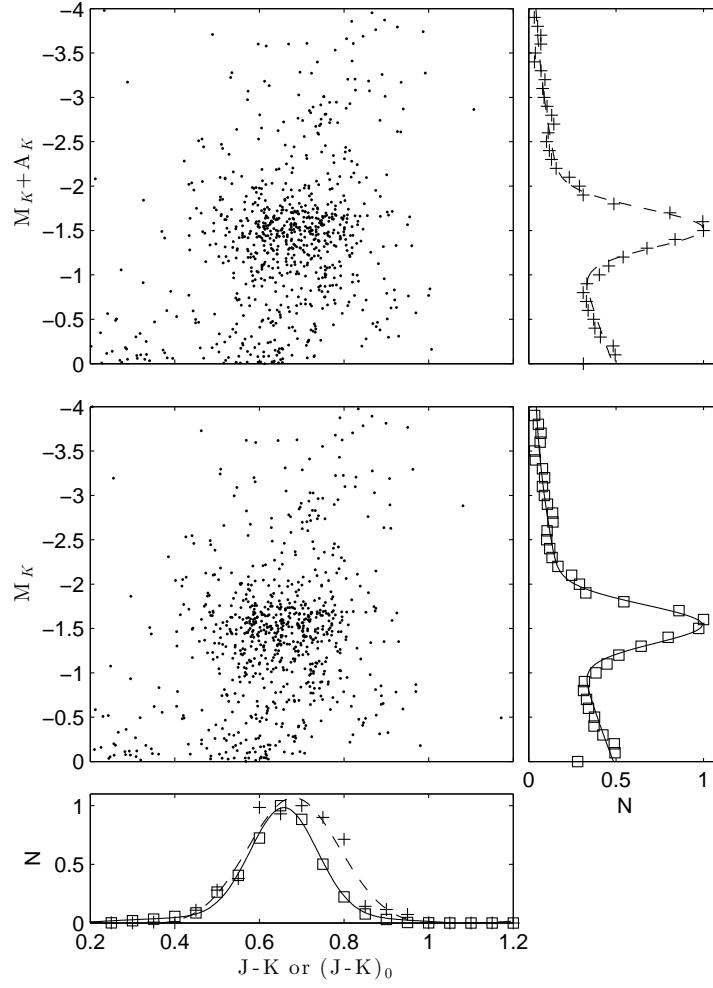


**Fig. 3** From top to bottom, the metallicity bins are  $(-0.6, -0.3)$ ,  $(-0.3, 0.0)$ , and  $(0.0, 0.4)$ , respectively. The left panels present the distributions of the full sample of the giant stars with different metallicity, respectively. The middle panels present the best fit distribution models of RGB stars. The right panels present the distributions of the residuals of full sample in the residuals of full sample after subtracting the RGB distributions in the middle panels. They are mostly contributed by the RC stars. The red and yellow contours show the 68% and 95% completeness of the RC candidates, respectively. The horizontal and vertical white lines give cuts for stars with  $\log g > 2.9$  dex,  $\log g < 2.1$  dex and  $T_{\text{eff}} > 5200$  K, respectively. The leaning white line located at  $\log g = 0.0016T_{\text{eff}} - 4.7170$  gives another cut for removing the bump stars located at the bottom-right corner at the 95% contour in the bottom-right panel. The color codes the stellar count in the bins.

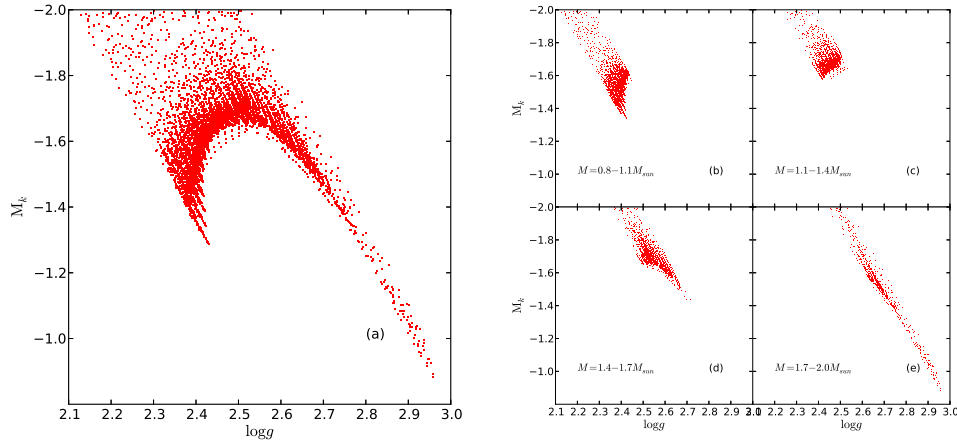


**Fig. 4** The colors show the fractions of true RC stars with different metallicity bins,  $(-0.6, -0.3)$ ,  $(-0.3, 0.0)$ , and  $(0.0, 0.4)$ , from left to right panel, respectively, in the  $\log g$  vs.  $T_{\text{eff}}$  plane.

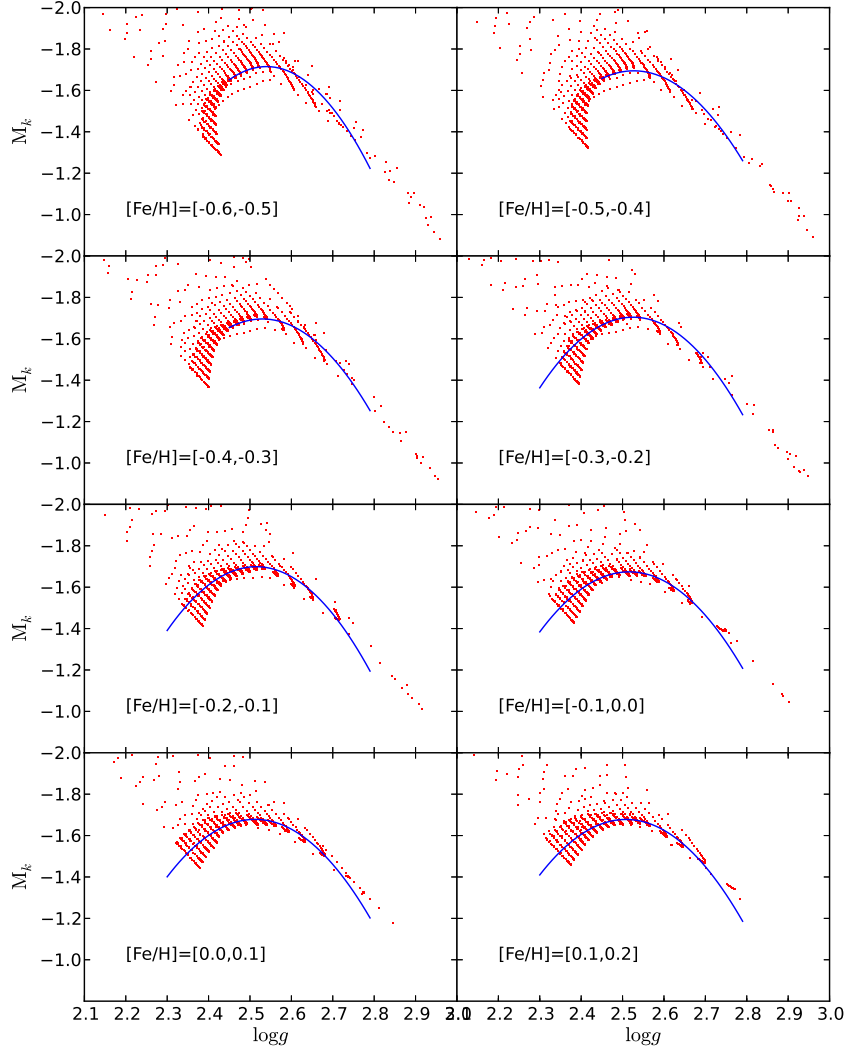




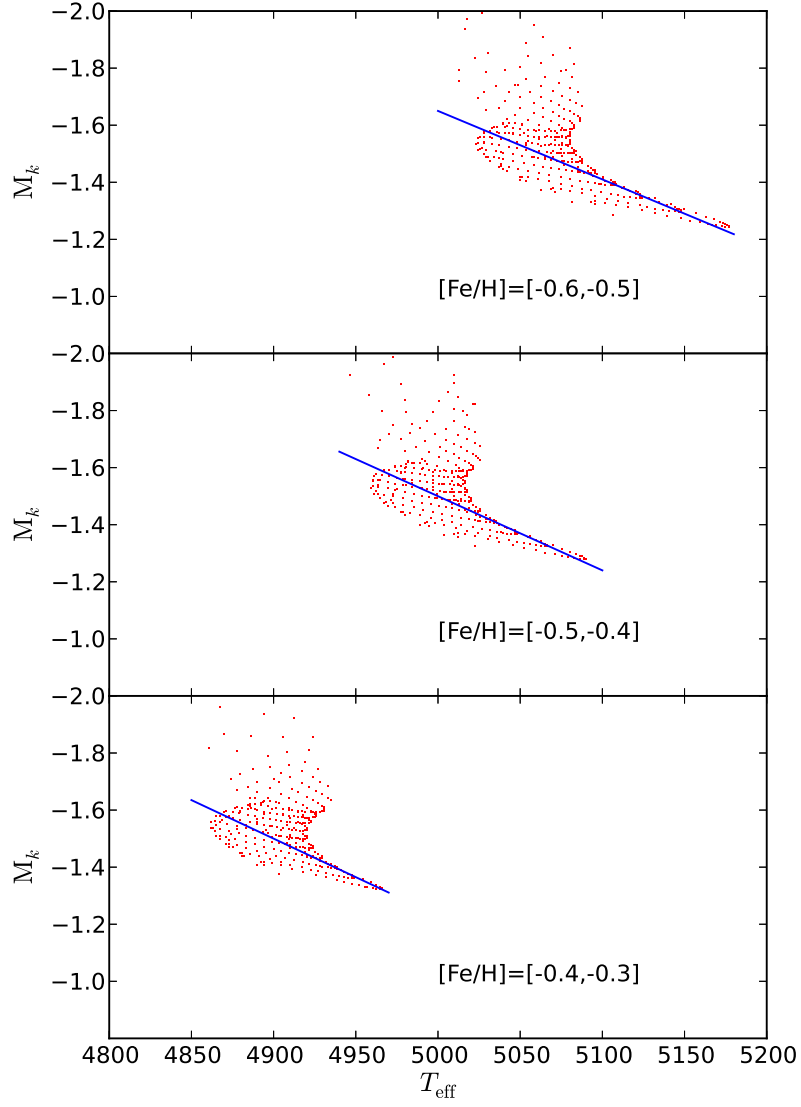
**Fig. 5** The top-left panel shows the reddened  $J - K$  vs.  $M_K + A_K$  diagram for about 900 giant stars selected from *Hipparcos* catalog. The middle-left panel shows the similar plot but with the dereddened color index and absolute magnitude. The bottom panel shows the marginalized distribution of the reddened  $J - K$  for the stars (corresponding to the top-left panel) with cross symbols and that of the dereddened color index  $(J - K)_0$  (corresponding to the middle-left panel) with the square symbols. The dashed and solid lines are the best fit model of Eq (7) for the  $J - K$  and  $(J - K)_0$ , respectively. The top-right panel shows the marginalized distribution of  $M_K + A_K$  with cross symbols. The dashed line stands for the best fit model according to Eq. (6). The middle-right panel shows the marginalized distribution of the dereddened  $M_K$  with square symbols, and the best fit model with the solid line.



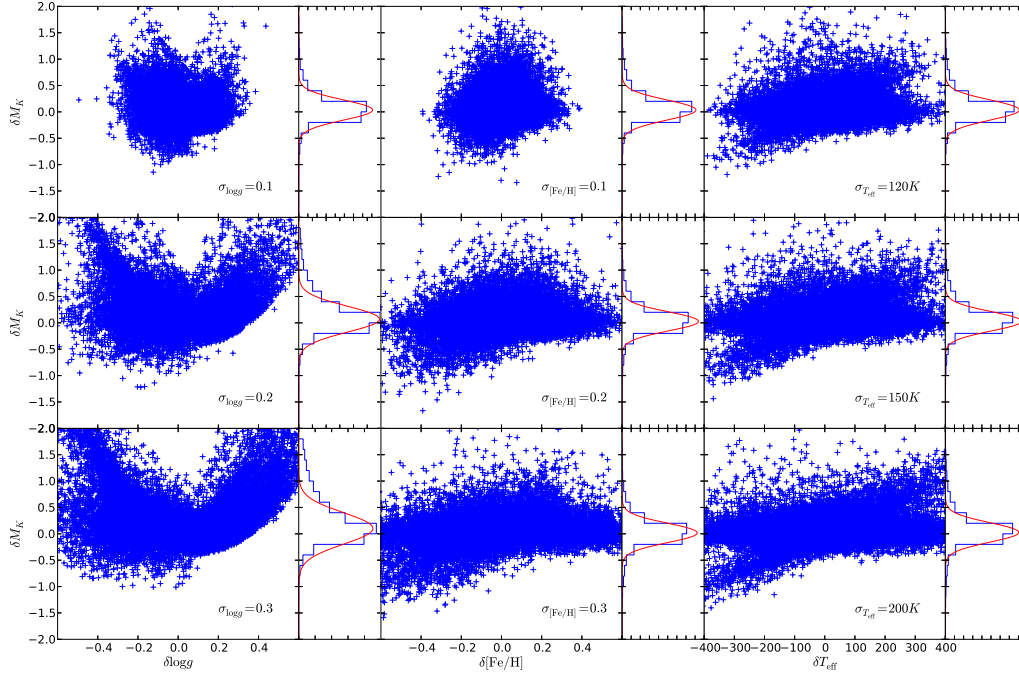
**Fig. 6** Panel (a): The synthetic  $\log g$  vs.  $M_K$  diagram for RC stars from PARSEC stellar evolution track (Bressan et al. 2012). The range of Age is from 4 Myr to 13 Gyr with steps of  $\Delta(\log t) = 0.05$ . The ranges of Metallicity [Fe/H] and  $M_{\text{ini}}$  are  $-0.6 \sim 0.3$  ( $Z_{\odot}=0.0152$ ) and  $0.8 \sim 2 M_{\odot}$  respectively. Panel (b)-(e): The  $\log g$ - $M_K$  relation for RC stars with  $M_{\text{ini}}$  between  $0.8 \sim 1.1 M_{\odot}$ ,  $1.1 \sim 1.4 M_{\odot}$ ,  $1.4 \sim 1.7 M_{\odot}$ , and  $1.7 \sim 2.0 M_{\odot}$ , respectively.



**Fig. 7** For RC stars with  $\log g > 2.45$ , the  $M_K$  is modeled as quadratic polynomials of  $\log g$  in each  $[\text{Fe}/\text{H}]$  bin. The red dots are the synthetic data and the blue lines are the best fit quadratic polynomials.

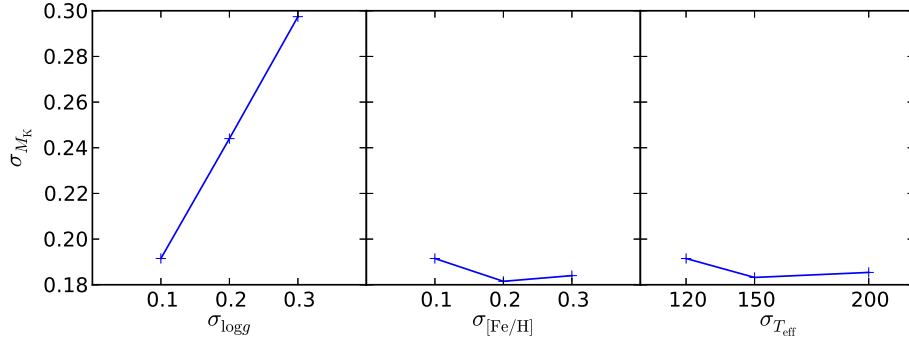


**Fig. 8** For RC stars with  $\log g < 2.45$  and  $[\text{Fe}/\text{H}] < -0.3$ , the  $M_K$  is modeled linearly as a function of  $T_{\text{eff}}$  in each  $[\text{Fe}/\text{H}]$  bin. The red dots are the synthetic data and the blue lines are the best fit lines.

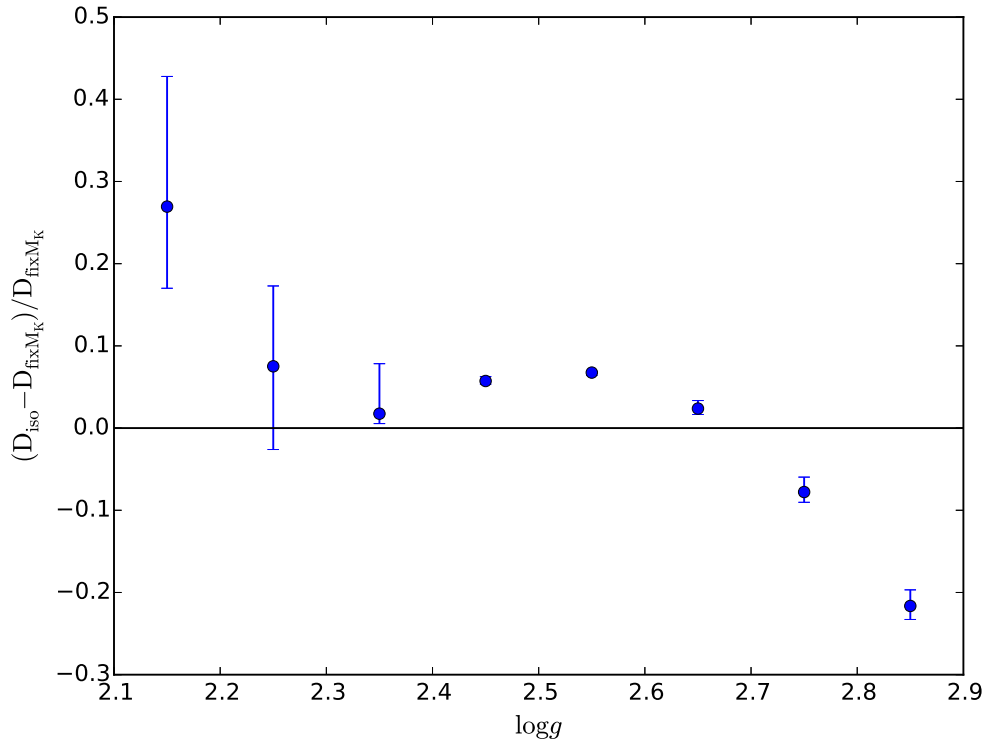


**Fig. 9** The scatter plot of the residuals of derived  $M_K$ , denoted by  $\delta M_K$ , for 20,000 simulate data with various uncertainties in  $\log g$  (left panels),  $[\text{Fe}/\text{H}]$  (middle panels), and  $T_{\text{eff}}$  (right panels).





**Fig. 10** The relationship between the standard deviation of the residual  $M_K$ , denoted by  $\sigma_{M_K}$ , and the uncertainty of  $\log g$  ( $\sigma_{\log g}$ ),  $[\text{Fe}/\text{H}]$  ( $\sigma_{[\text{Fe}/\text{H}]}$ ), and  $T_{\text{eff}}$  ( $\sigma_{T_{\text{eff}}}$ ).



**Fig. 11** The comparison between the the isochrone-based ( $D_{\text{iso}}$ ) and fixed absolute magnitude-based ( $D_{\text{fix}M_K}$ ) distance estimates at various  $\log g$ .